

Integrative RNA-seq and methylation analysis

AT Alexander Trostle LL Laura A Lavery YW Ying-Wooi Wan HYZ Huda Y Zoghbi

Updated date: May 4, 2021

An abbreviated version of this protocol was published in eLIFE in Mar 2020

Losing Dnmt3a dependent methylation in inhibitory neurons impairs neural function by a mechanism impacting Rett syndrome

DOI: 10.7554/eLife.52981

Detailed protocol

Thank you for your questions. Please find below the detailed methods as requested.

Estimating methylation levels

Methylation levels were calculated as percentages as cited in our methods from Sabbagh et al. In Sabbagh et al., they used "call_methylated_sites from Methylypy package (<https://github.com/yupenghe/methylpy>)" to determine methylation level at individual cytosines.

Pearson correlations in figure 2 were calculated with the cor function in R and run as below:

Fig2E. cor((vector of wild type mCH values in gene body), (vector of wild type mCG values in gene body), method = c("pearson"))

Fig2F. cor(((vector of wild type mCH values in gene body) - (vector of DKO mCH values in gene body)), ((vector of wild type mCG values in gene body) - (vector of DKO mCH values in gene body)), method = c("pearson"))

Integrative RNA-seq and methylation analysis

In Figure 4C, methylation versus gene expression plots were made using running average binning on significantly differentially expressed genes (DEGs, padj < 0.01). Below are the steps taken to generate these plots in more detail.

1. Methylation levels were determined as described above
2. Calculate change in % methylation as ((% mCH in cKO) - (% mCH in WT)) or as ((% mCG in cKO) - (% mCG in WT)). Here 'cKO' is either Dnmt3a or MeCP2 conditional knockout (cKO) and 'WT' is wildtype mice with endogenous levels of either protein.
3. Order genes based on ascending methylation value.
4. Bin so that each bin contains the same number of genes, with 80 percent gene overlap between consecutive bins.
 - a. For the plots of genes significantly misregulated only in the *Dnmt3a* cKO model, each bin has 25 genes and the window moves by five genes per bin. This results in 126 bins.
 - b. For plots of genes significantly misregulated only in the *Mecp2* cKO model, the bin size is 10 genes and the window slides by two each time. This results in 62 bins.
 - c. For plots of genes significantly misregulated in both models, the bin size is 10 genes and the window slides by two each time. This results in 42 bins.
5. Plot one point per bin.
6. After binning, a univariate linear model was fit to the data with the lm function in R, and the R^2 (percentage variance in log2 fold change explained by methylation) was calculated.
7. Compare these R^2 values to R^2 values from 1000 random gene repetitions, with each repetition containing the same number of non-differentially expressed genes (padj > 0.01).
8. Plot with gray points represent each iteration of this process, with the original DEG (padj < 0.01) values for R^2 shown in larger orange, green, or blue points.
9. Compute P value as $(r+1)/(n+1)$ as cited in our methods from North et al.; r is number of repetitions where R^2 is greater than that in the DEGs, and n is total number of repetitions.

How to cite: (Readers should cite both the Bio-protocol preprint and the original research article where this protocol was used)

1. Trostle, A. , Lavery, L. , Wan, Y. and Zoghbi, H. Y.(2021). Integrative RNA-seq and methylation analysis. Bio-protocol Preprint. [bio-protocol.org/prep1053](https://doi.org/10.21956/bio-protocol.d1053).
2. Lavery, L. A., Ure, K., Wan, Y., Luo, C., Trostle, A. J., Wang, W., Jin, H., Lopez, J., Lucero, J., Durham, M. A., Castanon, R., Nery, J. R., Liu, Z., Goodell, M., Ecker, J. R., Behrens, M. M. and Zoghbi, H. Y.(2020). Losing Dnmt3a dependent methylation in inhibitory neurons impairs neural function by a mechanism impacting Rett syndrome. eLIFE. DOI: [10.7554/eLife.52981](https://doi.org/10.7554/eLife.52981)

Copyright: Content may be subjected to copyright.